Joint-ViVo: Selecting and Weighting Visual Words Jointly for Bag-of-Features based Tissue Classification in Medical Images

Jingyan Wang[a,*]

[a]*Mathematical and Computer Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*

Abstract

Automatically classifying the tissues types of Region of Interest (ROI) in medical imaging has been an important application in Computer-Aided Diagnosis (CAD), such as classification of breast parenchymal tissue in the mammogram, classify lung disease patterns in High-Resolution Computed Tomography (HRCT) etc. Recently, bag-of-features method has shown its power in this field, treating each ROI as a set of local features. In this paper, we investigate using the bag-of-features strategy to classify the tissue types in medical imaging applications. Two important issues are considered here: the visual vocabulary learning and weighting. Although there are already plenty of algorithms to deal with them, all of them treat them independently, namely, the vocabulary learned first and then the histogram weighted. Inspired by Auto-Context who learns the features and classifier jointly, we try to develop a novel algorithm that learns the vocabulary and weights jointly. The new algorithm, called Joint-ViVo, works in an iterative way. In each iteration, we first learn the weights for each visual word by maximizing the margin of ROI triplets, and then select the most discriminate visual words based on the learned weights for the next iteration. We test our algorithm on three tissue classification tasks: identifying brain tissue type in magnetic resonance imaging (MRI), classifying lung tissue

*Corresponding author. Tel: 966-544701866
Email address: jingyan.wang@kaust.edu.sa (Jingyan Wang)

in HRCT images, and classifying breast tissue density in mammograms. The results show that Joint-ViVo can perform effectively for classifying tissues.
Keywords: Computer-Aided Diagnosis, Tissue Classification, Bag-of-Features, Visual Vocabulary, Visual Word Weighting

1. Introduction

Automated Computer Aided Diagnosis (CAD) systems are playing an important role in modern medical practices [1, 2, 3]. Accurate classification of medical images according to tissue type at the region of interest (ROI) level is important in many CAD applications [4]. A typical application is the diagnosis of breast cancer using mammogram as the medical imaging technology [2]. From a medical point of view, it is well-known that there is a strong positive correlation between high breast parenchymal density and high breast cancer risk. Thus, the development of automatic methods for classification of breast parenchymal tissue in the mammogram is justified for an automatic risk assessment framework in prospective CAD systems. Several techniques have been proposed for breast density classification using mammogram [5, 6, 7]. Another typical application is the diagnosis of diffuse lung diseases (DLDs), which are a heterogeneous group of diseases that affect the lung parenchyma in various ways [8]. High-resolution computed tomography (HRCT) [9] is useful to characterize DLDs because it provides better delineation of small structures and details within the lung. However, interpreting HRCT images is difficult even for specialists because of the complexity and variation in diffuse disease patterns. Therefore, CAD system to classify lung disease patterns is required. In recent years, many automated techniques have been proposed to classify diffuse lung disease patterns into several classes such as ground-grass opacities, reticular and linear opacities, honeycombing, emphysematous change, and so on [10].

Most of the tissue classification techniques extract features from medical images for classification with texture analysis approaches such as the gray level co-occurrence matrix (GLCM) [11] etc., which measure spatial dependencies of

intensity values within a region of interest (ROI) as second and higher statistics. Although they offer significant discriminatory power between various tissue patterns, the approaches do not work well for the patterns with inhomogeneous texture distribution within a ROI, such as the reticular patterns and the honeycombing patterns, because the statistics can only capture averaged feature over the ROI. To overcome this difficulty, the bag-of-features model is introduced for tissue classification of ROI in medical image, which have recently been proven to be effective for image retrieval task [12, 13, 14, 15]. Barnathan et, al. [16] proposed a methodology for discriminating between various types of normal and diseased brain tissue in medical images that utilizes bag-of-features, to extract discriminative texture features. Rather than focusing on images of the entire brain, they extracted local descriptors for individual ROI as determined by domain experts, and represented it as a frequency of codebook. Kato et, al. [10] proposed a bag-of-features approach for improvement of lung tissue classification in diffuse lung disease. In their model, images are represented as histograms or distributions of several types of local features that are obtained from training samples automatically. Bosch et, al. [7] presented a bag-of-features based approach to model and classify breast parenchymal tissue in mammogram, using a classifier based on local descriptors and probabilistic Latent Semantic Analysis (pLSA) [17], which a generative model from the statistical text literature.

As argued by Cai et, al. [18], the Visual Vocabulary (ViVo) (or codebook) plays the key role in the bag-of-feature model, and it is a collection of vector quantized features. The most popular way of creating visual vocabulary is by using k-means clustering [19, 20] or its variant, i.e., hierarchical k-means. However, it is argued that k-means does not select the most informative descriptors as it tends to concentrate the cluster centers in high density areas of the feature space and starves lower density ones [21, 22]. To deal with this problem, two kinds of strategies have been proposed:

- **Learning discriminative visual vocabulary**: In [21, 22], radius-based clustering is used for visual vocabulary (while) generation. In [23], su-

pervised learning of quantizer codebooks is proposed by information loss minimization.

- **Weighting visual words**: In [18], Cai et al. presented a visual word weighting strategy by learning a weighted similarity metric to satisfy that the weighted similarity between the same labeled images is larger than that between the differently labeled images with largest margin. In [13, 15], Wang et al. proposed a novel visual word weighting method by analyzing the discriminative power of each visual word by the sub-similarity function in the bin that corresponds to the visual word.

Up to now, the above two strategies are always used independently. First, the visual vocabulary is learned and then the weighting factors are estimated for each visual word. In this case, employing discriminative visual vocabulary generation methods without taking the visual words weighting into account is suboptimal. On the one hand, the bag level features is based on the histogram of the local features, which is directly determined by the visual vocabulary. While on the other hand, the visual word weighting is learned by the supervision of the labels of the features in [13, 18, 24] based on the bag-level features of ROI, and then a classifier will be trained for the weighted bag-of-features. Apparently, the weighting of the bag level feature has close relationship with the construction of the bag level feature itself, which is furthermore determined by the visual vocabulary. A better way is to learn the visual vocabulary and the weighting factors jointly. Nonetheless, it seems that the problem of deriving a joint procedure in the presence of combined clustering and weighting has received no attention.

Inspired by Auto-context [25, 26], which learn the classifier and features jointly in a iterative algorithm, we try to develop a novel joint visual vocabulary learning and visual words weighting estimation algorithm. In a transitional way for pattern classification, we first extract the features for samples and then train a classifier for these samples features. The feature extraction and classifier training is done inexpediently, assuming that the feature construction and the

classifier are not related. A break-through thought proposed by [25, 26] is that the features of samples can also be constructed by previous classification results in an iterative procedure, and thus the subsequent classifier will be trained using the newly constructed feature. The most important highlight of this algorithm is combining classification map $\mathcal{P}^{(t)}(i)$ and image patch $X(N_i)$ as features of pixel i for the training of classifier $F_k^{(t)}$. The classification map $\mathcal{P}^{(t)}(i)$ is determined by classifier $F_k^{(t-1)}$ in turn. On the other hand, the classifier $F_k^{(t)}$ will also be trained using $\mathcal{P}^{(t)}(i)$ as part of features. In this way, we can learn the feature and the classifier jointly in an iterative way. Now we take a look at the bag-of-features methods: the visual word weighting is learned for fixed bag level features, which are in turn constructed according to the visual vocabulary, and the vocabulary itself is learning using a clustering algorithm. Here, like Auto-Context, we try to guide the clustering procedure according to the weighting learned. That means we re-select the cluster center of the training local feature set according to an object function used to learn the visual word weights. We repeat this in a iterative algorithm until convergence. In this way, we can **joint**ly learn a discriminative **vi**sual **vo**cabulary and its corresponding visual word weights, so we call it **Joint-ViVo** algorithm. In contrast to many independent visual vocabulary learning methods proposed in [21, 22] and in contrast to the weighting factors estimation method proposed in [18, 24], the joint method we propose is neither a novel vocabulary learning algorithm nor does it assign weights to learned visual words. It jointly select local features from the training sets and assign them with weights iteratively. The learned visual vocabulary and its weighting should be used together to achieve an optimal performance.

In Section 2, we rigorously formulate the problem addressed in this paper by introducing the bag-of-features framework. In section 3, we introduce the proposed visual words selection and weighting algorithm— Joint-ViVo. Section 4 introduces the experimental methodology and reports experimental results and discussions. Section 5 concludes this paper with future work.

2. Bag-of-Features Based Tissue Classification Framework

Below, we present a bird's eye view of bag-of-features based tissue classification in medical images. A block diagram of bag-of-feature based tissue classification system is shown in Fig 1.

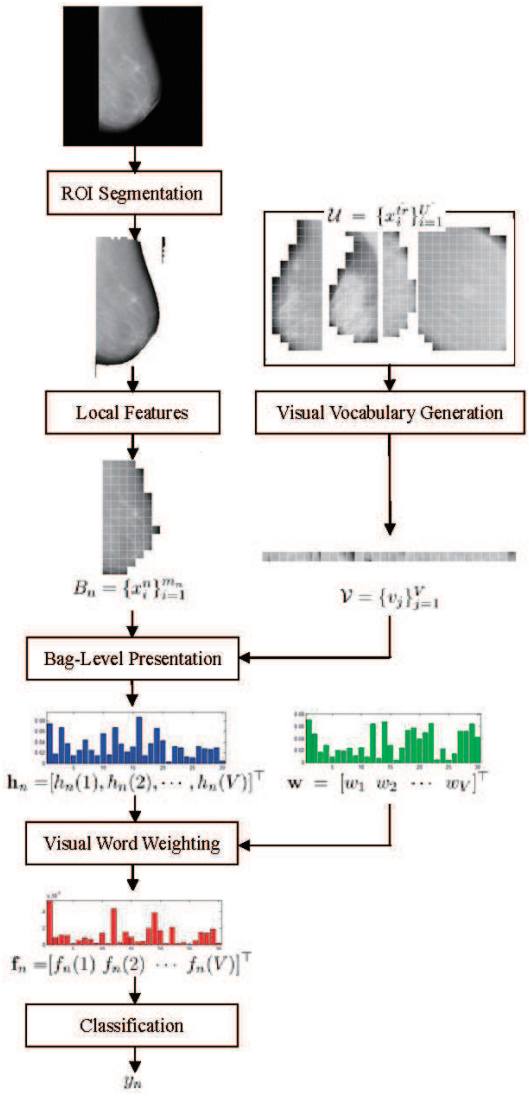


Figure 1: Block diagram of bag-of-feature based tissue classification system.

- **ROI Segmentation** The first stage of the system is to segment the medical image and extract the regain of interest (ROI). The obtained training ROI set is denoted as $\mathcal{N} = \{B_n, y_n\}_{n=1}^{N}$, where y_n is n-th ROI's class label. Varies segmentation ROI methods have been proposed. Since this is not the focus of this paper, we simply apply the existed segmentation methods [7] or segment ROI manually [10, 16].

- **Local Features** The next stage is to represent a ROI B_n as a collection of local features, such as image patches [27] (also called textons, intensity descriptor, or small block) and key points with SIFT descriptors [28]. We represent ROI B_n as a bag, which contains m_n local features denoted by $B_n = \{x_i^n\}_{i=1}^{m_n}$.

- **Visual Vocabulary Generation** To obtain a visual vocabulary $\mathcal{V} = \{v_j\}_{j=1}^{V}$ of size V, we usually apply a clustering algorithm on the training local feature sets $\mathcal{U} = \{x_i^{tr}\}_{i=1}^{U}$. The cluster centers $v_j, j = 1, \cdots, V$ will be used as visual words.

- **Bag-Level Presentation** We apply the *kernel codebook* proposed in [29] to represent a image ROI $B_n = \{x_i^n\}_{i=1}^{m_n}$ as a soft histogram of visual vocabulary $\mathcal{V}$, generating a V-dimensional frequency vector H_n

$$\begin{aligned} \mathbf{h}_n =& [h_n(1), h_n(2), \cdots, h_n(V)]^\top \\ h_n(j) =& \frac{1}{m_n} \sum_{i=1}^{m_n} K(x_i^n - v_j) \end{aligned} \tag{1}$$

  where $K(x_i^n - v_k) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{||x_i^n - v_k||^2}{2\sigma^2}\right)$ is Gaussian-shaped kernel and σ is the smoothing parameter of kernel K.

- **Visual Word Weighting** An important procedure in bag-of-feature based image retrieval and classification is to weight the histogram vectors according to its discriminant ability:

$$\begin{aligned} f_n(j) =& w_j \cdot h_n(j) \\ \mathbf{f}_n =& [f_n(1)\ f_n(2)\ \cdots\ f_n(V)]^\top \end{aligned} \tag{2}$$

where w_j is the weight for j-th visual word.

- **Classification** After each ROI in a medical image is represented as a bag-level feature vector $\mathbf{f}_n$, we can train a classifier to distinguish the ROI according to its tissue types.

3. Joint-ViVo: Joint Learning and Weighting of Visual Words

In this section, we will firstly discuss the weighting of visual words given the visual vocabulary, using an object function designed to distinguish a triplet of ROIs. Then we give a novel visual words updating methods according to the learned weighting to minimize the object function. Finally, we combine these tow procedure and obtain an iterative visual words selection and weighting algorithm, called Joint-ViVo.

3.1. Discriminative Visual Word weighting

3.1.1. Bag-level Similarity and Distance Vector

The histogram intersection kernel $s(\mathbf{h}_p, \mathbf{h}_q) = \sum_j min(h_p(j), h_q(j))$ is commonly used to compute the similarity between a pair of bag-level feature [30]. The intersection of the j-th visual word occurrence frequency between bag B_p and bag B_q is denoted by $s_{pq}(j) = min(h_p(j), h_q(j))$ in the bag-of-features model. Accordingly, $\mathbf{s}_{pq}$ represents the intersection vector between bag B_p and bag B_q:

$$\mathbf{s}_{pq} = [s_{pq}(1)\ s_{pq}(2)\ \cdots\ s_{pq}(V)]^\top \tag{3}$$

In the typical bag-of-features model, the similarity between two bags is the sum of the equally weighted intersections: $s(B_p, B_q) = \sum_j s_{pq}(j)$ [18]. In contrast, we assign different weights for visual words, resulting in a weighted similarity defined as $s_{\mathbf{w}}(B_p, B_q) = \mathbf{w}^\top \mathbf{s}_{pq} = \sum_j w_j s_{pq}(j)$, which is a similarity metric according to the definition in [18].

Instead of measure the similarity between two bags, we can also compute the distance between two normalized frequency histograms using the χ^2 statistic

[10, 31], where the j-th visual word χ^2 distance between bags B_p and bag B_q is j-th $d_{pq}(j) = \frac{1}{2}\frac{(h_p(j)-h_q(j))^2}{h_p(j)+h_q(j)}$. The χ^2 distance vector between bags B_p and bag B_q is denoted as

$$\mathbf{d}_{pq} = [d_{pq}(1)\ d_{pq}(2)\ \cdots\ d_{pq}(V)]^\top \tag{4}$$

In this way, we compute the χ^2 distance between bags B_p and bag B_q as $d(B_p, B_q) = \sum_j d_{pq}(j)$. Similarly, we assign different weights for visual words, resulting in a weighted χ^2 distance defined as $d_{\mathbf{w}}(B_p, B_q) = \mathbf{w}^\top \mathbf{d}_{pq} = \sum_j w_j d_{pq}(j) = \frac{1}{2}\sum_j w_j \frac{(h_p(j)-h_q(j))^2}{h_p(j)+h_q(j)}$. We must note that, instead of weighting the histogram as $f(j) = w_j \times h(j)$ directly in (2), we weight the j-th visual word χ^2 distance $d_{pq}(j)$.

3.1.2. Large Margin Based Weights Vector $\boldsymbol{w}$ Learning

Inspired by BoostMap [32], we learn our visual word weighting vector $\mathbf{w} = [w_1\ w_2\ \cdots\ w_V]^\top$ by classifying the triplets of objects in the dataset. This methodology are also used in [18, 33, 34]. Let $\mathcal{T}$ be a triplet index set of training ROIs presented as bags of features: $\mathcal{T} = \{(n, p, q) | y_n = y_p\ and\ y_n \neq y_q\}$, where y_n denotes the class label for bag B_n. We aim to make the weighted similarity between same labeled images larger than that between differently labeled images. Ideally, the learnt weight vector $\mathbf{w} \in \mathbb{R}_+^V$ satisfies the constraint

$$\mathbf{w}^\top \mathbf{s}_{nq} < \mathbf{w}^\top \mathbf{s}_{np}, \forall (n, p, q) \in \mathcal{T} \tag{5}$$

when the distance measure is used, we have

$$\mathbf{w}^\top \mathbf{d}_{np} < \mathbf{w}^\top \mathbf{d}_{nq}, \forall (n, p, q) \in \mathcal{T} \tag{6}$$

The margin of triplet $\phi = (n, p, q)$ with respect to $\mathbf{w}$, is then computed as

$$\begin{aligned} \rho_\phi(\mathbf{w}) &= \mathbf{w}^\top \mathbf{s}_{nq} - \mathbf{w}^\top \mathbf{s}_{np} \\ &= \mathbf{w}^\top \times \mathbf{z}_n \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathbf{z}_\phi &= \mathbf{s}_{nq} - \mathbf{s}_{np} \\ &= [s_{nq}(1) - s_{np}(1), \cdots, s_{nq}(V) - s_{np}(V)]^\top \\ s_{nq}(j) &= min(h_n(j), h_q(j)) \end{aligned} \tag{8}$$

is the original margin vector. Similarly, a $\mathbf{z}_\phi = \mathbf{d}_{np} - \mathbf{d}_{nq}$ can also be defined using distance measure instead of similarity.

After the margins are defined, the problem of learning feature weights $\mathbf{w}$ can be solved within the large margin framework. We perform the estimation in a popular margin formulations— the logistic regression formulation. We also add an L_1 norm penalty of $\mathbf{w}$ to the objective function to encourage the sparseness of weight, which leads to the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}} \ Q(\mathbf{w}) = \sum_{\phi} log(1 + exp(-\mathbf{w}^\top \times \mathbf{z}_\phi)) + \lambda ||\mathbf{w}||_1; \\ &s.t \ \mathbf{w} \geq 0. \end{aligned} \tag{9}$$

For fixed $\mathbf{z}_\phi$, (9) is a constrained convex optimization problem. Due to the nonnegative constraint on $\mathbf{w}$, it cannot be solved directly by using gradient descent. To overcome this difficulty, we set $w_j = u_j^2, j = 1, \cdots, V$ and reformulate the problem slightly as:

$$\min_{\mathbf{u}} \ O(\mathbf{u}) = \sum_{\phi} log(1 + exp(-\sum_{j=1}^{V} u_j^2 \mathbf{z}_\phi(j)) + \lambda ||\mathbf{u}||_2^2 \tag{10}$$

thus obtaining an unconstrained optimization problem. The solution of $\mathbf{u}$ can thus be readily found through gradient descent with a simple update rule:

$$\mathbf{u} \leftarrow \mathbf{u} - \eta \nabla O(\mathbf{u}) \tag{11}$$

where $\nabla O(\mathbf{u}) = \left(\lambda \mathbf{1} - \sum_\phi \frac{exp(-\sum_j u_j^2 z_\phi(j))}{1+exp(-\sum_j u_j^2 z_\phi(j))} \mathbf{z}_\phi\right) \otimes \mathbf{u}$, $\otimes$ is the Hadamard product operator, η is the learning rate.

3.2. Visual Words Selection

After obtaining the visual words weighting vector $\mathbf{w}$ for bag-level histogram features, we can validate the selected visual words and update them accordingly. The visual words selection and the training local features clustering are done alternately. Given the initial visual words $\mathcal{V} = \{v_j\}_{j=1}^V$, each feature x_i^{tr} in $\mathcal{U}$ will be clustered to V clusters $\{C_j\}_{j=1}^V$ with $\{v_j\}_{j=1}^V$ as centroid as

$$C_j = \{x_i^{tr} | v_j = \underset{v_k \in \mathcal{V}}{argmin} \ ||v_k - x_i^{tr}||^2\} \tag{12}$$

We now can construct the bag-level features using the selected visual words in $\mathcal{V}$. We denote the bag-level features (histograms) as function of visual words $v_1, \cdots, v_V$ as

$$\mathbf{h}_n(\mathcal{V}) = [h_n(1|\mathcal{V}) \ \cdots \ h_n(V|\mathcal{V})]^\top \tag{13}$$

where

$$\begin{aligned} h_n(j|\mathcal{V}) &= h_n(j|v_1, \cdots, v_V) \\ &= \frac{1}{m_n} \sum_{i=1}^{m_n} K(x_i^n - v_j) \end{aligned} \tag{14}$$

Obviously, the j-th bin of histogram is the function of j-th visual word instead of the entire visual vocabulary.

Up to new, each bag can be represented as a histogram of the selected visual words. With these visual words we can also compute the margin vector for each triplet of bags $\phi = (n, p, q)$ as $\mathbf{z}_\phi(v_1, \cdots, v_V)$ using (3) or (4).

We now define the optimisation problem underlying the iterative framework in terms of the visual words $v_1, \cdots, v_V$, and weighting vector $\mathbf{w}$ as follows

$$\begin{aligned} &\min_{\mathbf{w}, v_1, \cdots, v_V} Q(\mathbf{w}, v_1, \cdots, v_V) \\ &\qquad = \sum_{\phi} log(1 + exp(-\mathbf{w}^\top \times \mathbf{z}_\phi(v_1, \cdots, v_V))) \\ &\qquad\quad + \lambda ||\mathbf{w}||_1; \\ &\quad s.t.\ \mathbf{w} \geq 0,\ v_j \in C_j,\ j = 1, \cdots, V. \end{aligned} \tag{15}$$

where $Q(\mathbf{w}, v_1, \cdots, v_V)$ is the same loss function as defined in (9), and $\mathbf{z}_\phi(v_1, \cdots, v_V)$ specifies the bag-level feature margin for the ϕ-th bag triplet given the indices of the selected visual words $v_1, \cdots, v_V$. Given the learned visual words weighting vector $\mathbf{w}$, we can further update the visual words for each cluster C_j. This is equivalent to minimizing $Q(\mathbf{w}, v_1, \cdots, v_V)$ in (15) with respect to $v_1, \cdots, v_V$. We adopt a reminiscent procedure of coordinate descent so as to update v_j for

each cluster as follows

$$
\begin{aligned}
v_1^{(t+1)} &= \underset{x_i^{tr} \in C_1^{(t)}}{argminQ}(\mathbf{w}, x_i^{tr}, v_2^{(t)}, \cdots, v_V^{(t)}) \\
&\cdots \\
v_j^{(t+1)} &= \underset{x_i^{tr} \in C_j^{(t)}}{argminQ}(\mathbf{w}, \cdots, v_{j-1}^{(t)}, x_i^{tr}, v_{j+1}^{(t)}, \cdots) \\
&\cdots \\
v_V^{(t+1)} &= \underset{x_i^{tr} \in C_V^{(t)}}{argminQ}(\mathbf{w}, v_1^{(t)}, \cdots, v_{V-1}^{(t)}, x_i^{tr})
\end{aligned}
\tag{16}
$$

where $v_j^{(t)}$ and $v_j^{(t+1)}$ correspond to the old and updated visual word for the j-th cluster respectively, $\mathbf{w}$ is fixed to $\mathbf{w}^{(t)}$ for the t-th iteration. Since $\mathbf{w}$ is unchanged during the update of γ, only the $\mathbf{z}_\phi(v_1, \cdots, v_V)$ term of (15), i.e. the margin term of the logistic regression objective function, is considered here. Each v_j is updated while fixing all other visual words.

After the new visual words are selected by updating $\mathcal{V}^{(t)} = \{v_j^{(t)}\}_{j=1}^V$, we can then re-cluster the training local features $\mathcal{U}$ by assigning each local feature x_i^{tr} to its nearest visual word as (12) and the updated clusters $\{C_j^{(t)}\}_{j=1}^V$ are obtained. Then, using the new selected visual words, each bag B_n can be represented as a frequency vector $\mathbf{h}_n(\mathcal{V}^{(t)})$ and the weighting vector $\mathbf{w}$ can be updated by (9).

3.3. Joint Visual Words Selection and Weighting–Concluding Results

We summarize the traditional independent and our joint visual vocabulary and weight learning algorithms in Fig. 2. As we can see from Fig. 2 (a), the visual vocabulary $\mathcal{V}$ is firstly learned using a clustering algorithm, and then the each bag is represented as a histogram of $\mathcal{V}$. After this, the bag-level features are weighted resulting the visual words' weighting vector $\mathbf{w}$. Our Joint-ViVo in Fig. 2 (a) is basically similar to independent learning: using a visual vocabulary to build histogram as bag-level features and weight the features as visual words' weight. However, different from independent learning framework, the Joint-ViVo learns the weights and re-select the visual words from training local feature set alternately in a iterative procedure. Moreover, the re-clustering of the whole

training $\mathcal{U}$ is done according to the newly selected visual words $\mathcal{V}$, which play role of centroids, while the selection of visual words $\mathcal{V}$ is based on the prevenient learned weight vector $\mathbf{w}$. At the same time, the learning of the weight vector $\mathbf{w}$ is based on the bag-level features, which is the histogram of selected visual words $\mathcal{V}$ of a ROI. Thus, this enable an iterative algorithm as shown in Fig. 2 (b). We give the novel developed algorithm Joint-ViVo in algorithm 1.

Algorithm 1 Joint Learning and Weighting of Visual Words Algorithm: Joint-ViVo

Require: Local feature trainning set $\mathcal{U}$;

Require: Training bag triplet set $\mathcal{T}$;

Require: Initial visual vocabulary $\mathcal{V}^{(0)}$;

Require: Initial visual word weighting vector $\mathbf{w}^{(0)}$;

Require: Stop criterion θ.

Cluster local features in $\mathcal{U}$ to initial visual words in $\mathcal{V}^{(0)}$ and obtain initial clusters $\{C_j^{(0)}\}_{j=1}^{V}$ as in (12).

for $t = 1, \cdots, T$ **do**

Represent each bag B_n in each bag triplet as a bag-level feature $\mathbf{h}_n^{(t)} = \mathbf{h}_n(\mathcal{V}^{(t-1)})$ based on previous selected visual vocabulary using (13) and (14);

Compute original margin vector $\mathbf{z}_\phi^{(t)}$ for each triplet $\phi = (n, p, q)$ using (8);

Update the visual word weighting vector $\mathbf{w}^{(t)}$ by updating $\mathbf{u}^{(t)}$ as (11);

if $||\mathbf{w}^{(t)} - \mathbf{w}^{(t-1)}||^2 < \theta$ **then**

Break.

else

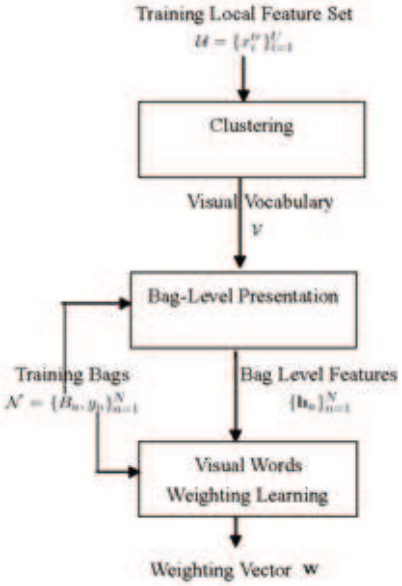
Update the visual words $\{v_j^{(t)}\}_{j=1}^{V}$ as in (16);

Re-cluster the training local set and obtain $\{C_j^{(t)}\}_{j=1}^{V}$ using (12).

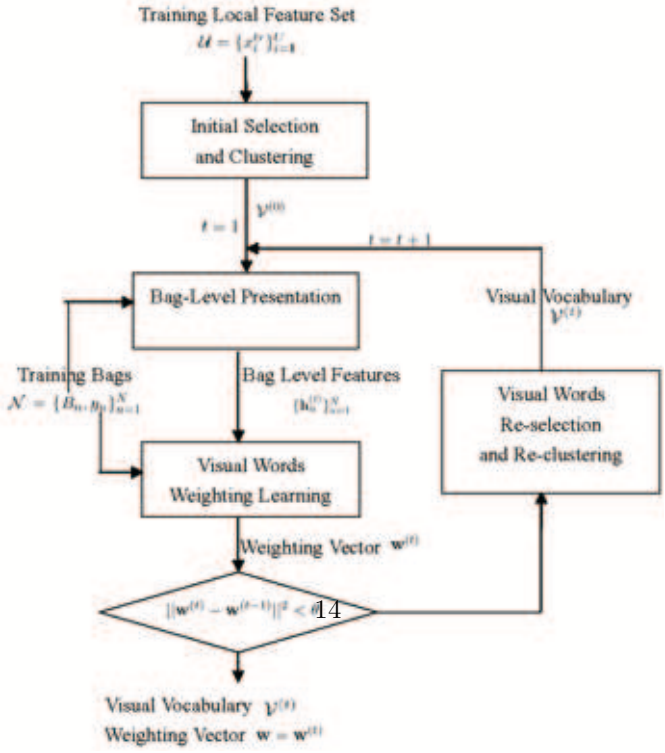
end if

end for

Output: Visual words $\mathcal{V}^{(t)}$ and weighting vector $\mathbf{w}^{(t)}$.


Training Local Feature Set
Clustering
Visual Vocabulary
Bag-Level Presentation
Training Bags
Bag Level Features
Visual Words
Weighting Learning
Weighting Vector

(a)


Training Local Feature Set
Initial Selection
and Clustering
Bag-Level Presentation
Visual Vocabulary
Training Bags
Bag Level Features
Visual Words
Re-selection
and Re-clustering
Visual Words
Weighting Learning
Weighting Vector
Visual Vocabulary
Weighting Vector

(b)

4. Experiments

In this section, we present the application of the Joint-VoVi on three tissue classification tasks:

1. Classifying Breast Tissue Density in Mammograms [7],
2. Classifying Lung Tissue in HRCT images [10].
3. Identifying Brain Tissue Type in MRI images [35],

In these three tasks, the system uses a nearly identical parameter setting, including the number of visual words and the stopping criterion.

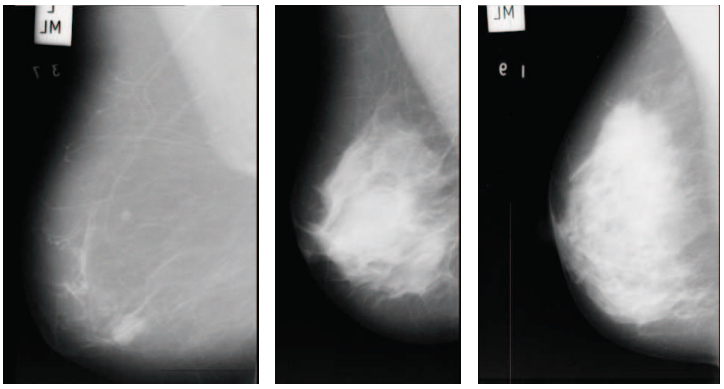
4.1. Experiment I: Classifying Breast Tissue Density in Mammogram

4.1.1. Mammogram Dataset and Setup

In this group of experiments, we test our method on classifying breast parenchymal tissue in Mammogram using bag-of-features model. The experiments are carried on a public and widely known database — Mammographic Image Analysis Society (MIAS) database [36]. This database is composed by the Medio- Lateral Oblique views of both breasts of 161 women (322 mammographies). The MIAS database provides annotations for each mammogram, and one of them is referred to the breast density. The images are labelled as (see Fig. 3):

1. Fatty (106 images): the breast is almost entirely fatty,
2. Glandular (104 images): the breast contains some fibroglandular tissue, or
3. Dense (112 images): the breast is extremely dense.

Given the set of training images, local descriptors are computed around the pixels of the tissue and a visual vocabulary $\mathcal{V}$ is obtained. We chose image patches [31] as the local feature for bag-of-features based classification breast tissue density in mammogram. A $N \times N$ square neighborhood is opened around each pixel. The pixels are row reordered to form a vector in an N^2 dimensional feature space. The patches are spaced by M pixels on a regular grid over the



(a) Fatty (b) Glandular (c) Dense

Figure 3: Three MIAS images belonging to one of each MIAS category.

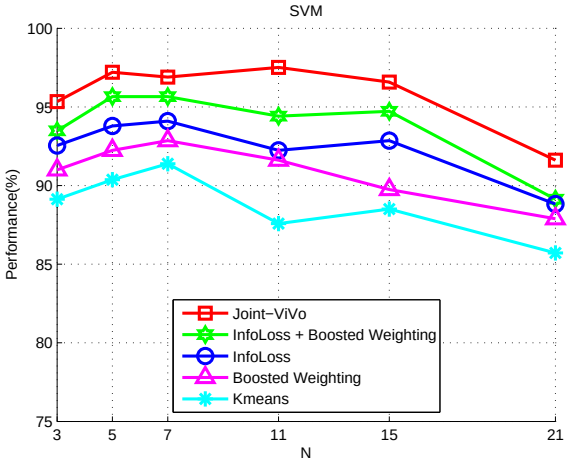
area of the tissue. The visual vocabulary $\mathcal{V}$ with its weighting factor vector $\mathbf{w}$ is learned using our Joint-ViVo algorithm. We use Kmeans [19, 20] as a baseline method to learn the visual vocabulary $\mathcal{V}$, and equal weighting as a baseline weighting vector $\mathbf{w}$. We also compare our Joint-ViVo algorithm to a state-of-the-art vocabulary learning algorithm — InfoLoss [23] and a state-of-the-art weighting algorithm — Boosted Weighting [13, 15]. After each ROI in mammogram is represented as a bag-level features using $\mathcal{V}$ and $\mathbf{w}$, we further perform the mammogram classification using SVM [37] and kNN [38].

In order to evaluate the results, we used a **leave-one-out** [39] method, in which each sample is analyzed by a classifier which is trained using all other samples. However when working with the MIAS dataset, we leave the two images (left and right breast) from the same woman. Therefore for the MIAS database we use 320 training images and 2 for testing 161 times, changing the test and train images every time. When using the SVM a Gaussian kernel is used, and the multi-class classification is done using the one-versus-all rule [40]. Overall performance rates are measured by the average value of the diagonal entries of the confusion table.

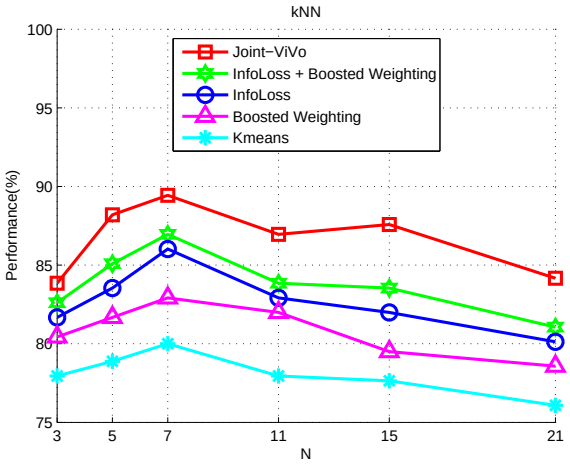
4.1.2. Results

For evaluation, a performance curve for each visual vocabulary learning or weighting method is plotted in Fig. 4 showing the classification rates versus N, which is the size of the texton, while fixing $V = 1600$ and $M = 2$. The other parameters of Joint-ViVo are selected by 10-fold cross validation on the training data set. From the results, we have the following observations:

1. Among InfoLoss, Boosted Weighting, their combination and Joint-ViVo, the proposed Joint-ViVo, gain the best performances in terms of average classification rate over all the 6 sizes of N. From Fig. 4 (a) using SVM, as for Joint-ViVo, it improves InfoLoss by 5.28% and Boosted Weighting by 6.83%. Furthermore, The combination of InfoLoss and Boosted Weighting improves InfoLoss by 2.17% and Boosted Weighting by 4.96%. Meanwhile, of all 6 texton sizes, the proposed Joint-ViVo perform best on $N = 11$. On the remaining texton sizes, their performances only slightly deteriorate compared to the other algorithms. Similar phenomena can also be observed in Fig. 4 (b). However, we must notice that the classification results of SVM is much better than kNN, which is not surprising.
2. From the results, it is clear that supervised vocabulary learning algorithms, especially Joint-ViVo and InfoLoss, outperform unsupervised Kmeans. Among the supervised vocabulary based methods, both the proposed Joint-ViVo and InfoLoss demonstrate excellent accuracy. It also shows that, on average, Boosted Weighting works better than the equal weighting method, but it can be further improved by using an effective vocabulary.
3. According to the results, we also arrived at the conclusion that Joint-ViVo have better performance than all independent vocabulary learning and weighting methods (or the combination of them). This is because, after jointly learning the visual vocabulary and its weighting vector, all discriminative information is contained exactly by the maximizing margins of ROI triplets.



(a) SVM



(b) kNN

Figure 4: Performances on MIAS dataset when changing the values of parameter N.

We also compare the obtained results on this database with those obtained by Blot and Zwiggelaar [41], Oliver et al. [42] and Anna Bosch et al. [7]. The experiment results are shown in Table 1.

Table 1: Comparison summary of the proposed method with other works that classify parenchymal density on MIAS database.

Reference	Performance (%)
[41]	50
[42]	73
[7]	91.39
Joint-ViVo	**97.52**

We did not use all the possible ROI triplets for weighting factor learning on the MIAS and DDSM data sets as the training data size is large and it takes hours to learn a visual vocabulary and weighting vector. However, compared to state-of-the-art methods, the classification result by using Joint-ViVo is already quite competitive as only the bag-of-features based methods [7] has similar performances. We attribute this result to the discriminate property of the local features in the data set, which is utilize by Joint-ViVo effectively. On all the three groups of experiments, the classification of Joint-ViVo with different annotations have better performances than other classifiers. These results suggest the effectiveness of Joint-ViVo compared with the representative methods for discriminative tissue classification.

4.2. Experiment II: Lung Tissue Classification in Diffuse Lung Disease

4.2.1. HRCT axial images dataset and experiment setup

In the second experiment, we classify the HRCT axial images with normal, emphysema, ground glass, honeycombing, and reticular patterns collected from a Hospital. A radiologist marked several ROIs by drawing boundaries that included abnormal patterns on each image. Obtained database are consists of 174 normal, 209 honeycombing, 346 emphysema, 189 reticular and 198 ground glass patterns.

For local feature extraction, the intensity and SIFT local features are calculated in 15×15 and 12×12 regions, respectively. The local features are sampled uniformly by sliding the local region with $M = 1, 3, 5$ pixel step. Joint-ViVo is

then employed to learn the vocabulary $\mathcal{V}$ and weights $\mathbf{w}$ from local feature set extracted from training images. The visual vocabulary size V is set to 1500. Then a ROI is characterized by a histogram of quantized local features, which are sampled from the ROI and quantized by using the visual vocabulary. The weighted histogram of the intensity or SIFT local features are used as a 1500-dimensional vector that is classified by the classifier. We use the SVM [37] as the classifier to classify the histogram features to 1 of the 4 diffuse lung diseases. We use the χ^2 kernel as the kernel function, which is extended Gaussian kernel with $\mathbf{w}$ weighted χ^2 distance as

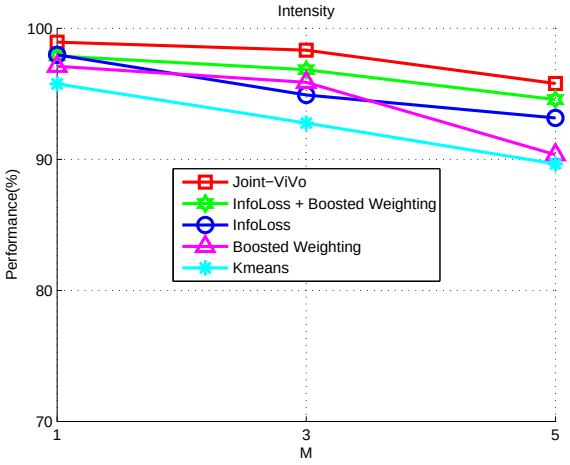
$$K(\mathbf{h}_p, \mathbf{h}_q) = exp\left(-\frac{1}{2g}\sum_{j=1}^{V} w_j \frac{(h_p(j)-h_q(j))^2}{h_p(j)+h_q(j)}\right) \tag{17}$$

For the multi-class classification, several SVM models are built using one versus one combinations, and classification is done by voting of these SVM models. The SVM parameters are optimized by the 2-fold cross validation. Classification performance is also evaluated by 2-fold cross validation.

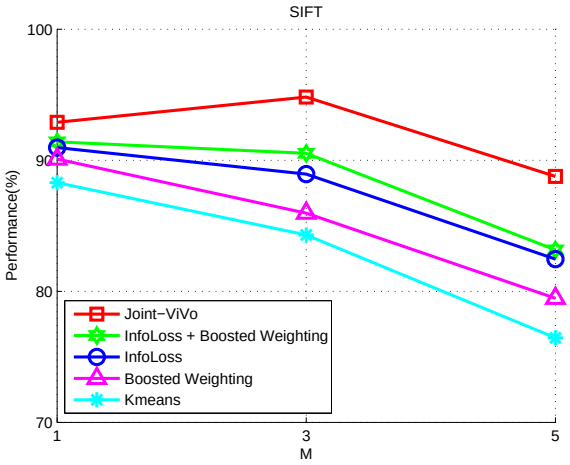
4.2.2. Results

Fig. 5 shows overall classification accuracy of the intensity and the SIFT feature with varying feature sampling condition. Fig. 5 shows the accuracies of the different methods using the ratio of the number of correctly classified ROI to the total number of ROI. The ratio is averaged over 2 trails for each method.

In the figure, it is clear that the k-means learned vocabulary with SIFT local feature gives the worst results. This is because the k-means simply find the vocabulary words according to the clustering centroid of the local feature set and this set is susceptible to noise has little discriminate information. The k-means learned vocabulary with intensity local feature perform better due to the intensity's robustness to the noise. Moreover, again, Joint-ViVo outperforms all other visual vocabulary learning and weighting strategies (including their combination). The Joint-ViVo method effectively rejects noise and outliers by selecting the most reliable visual words. Joint-ViVo learned vocabulary gives the best performance on all the cases. The overall recognition performance of Joint-



(a) Intensity



(b) SIFT

Figure 5: Overall classification accuracy of the intensity feature and the SIFT feature with different visual vocabulary learning methods.

ViVo appears to be very competitive. This is a strong evidence of the inherent relationship between vocabulary and its weights, through not clear yet, by can

be explored and employed by learning jointly. Moreover, The accuracy of both features increases as the number of the sampling points M. This means that sufficient number of local features are necessary to calculate smooth distribution of the features.

4.3. Experiment III: Identifying Brain Tissue Type in MRI

4.3.1. T1 and T2 MRI Dataset

In the third experiment we focus on classification of individual regions of interest in a dataset of 24 post-T1 weighted gadolinium-enhanced MRI slices and 96 pre-T2 weighted MRI slices of the brain of a single mouse. Images were registered prior to segmentation and normalized following combination. Segmentation itself was performed by a domain expert and supported by histology data. In particular, we are interested in discriminating between 21 T1 and 75 T2 manually segmented ROIs from these images, representing various types of tissue: cerebrospinal fluid (CSF), gray matter, tissue necrosis, hippocampus tissue, and samples from three distinct regions of tumor with varying degrees of vascularization, neoplastic growth, and tissue necrosis. We wish to discriminate between individual tissue types as well as collectively classify tissue as normal (CSF, hippocampus, and gray matter classes) or abnormal (necrosis, tumor1, tumor2, tumor3). The post-T1 ROIs were assigned labels "CSF", "Tumor1", "Tumor2", and "Tumor3", representing areas of cerebrospinal fluid, homogenous "typical" tumor tissue, heavily vascularized tumor tissue, and tumor tissue near an area of necrosis and edema. To take advantage of the imaging properties of T2 relaxation, we selected ROIs from the T2 image dataset in the following classes: "CSF", "Graymatter", "Hippocampus", and "Necrosis". These labels corresponded to areas of cerebrospinal fluid, normal gray matter tissue, a region of normal tissue located in the hippocampus, and a region of liquefactive necrosis near the lower central region of the tumor, respectively.

For local feature representation, given a fixed block size, each image is decomposed into a number of small blocks (image patches). Based on such small blocks from different images, a visual vocabulary $\mathcal{V}$ containing visual words (key

blocks) is generated using different algorithms including Joint-ViVo, InfoLoss [23] and Generalized Lloyd Algorithm (GLA) [43]. The visual weights $\mathbf{w}$ are also learned by Joint-ViVo and Boosted Weighting [13]. As a comparison, the Generalized Lloyd Algorithm (GLA) [43] and InfoLoss [23] are also used to generate the visual vocabulary $\mathcal{V}$, while Boosted Weighting [13] is used to learn the weighting of visual words. The visual word frequency vector (or weighted visual word frequency vector when visual words weighting is applied) is used as a representative feature vector of the image texture in classification. We then employ the Histogram Model, which has been shown effective for texture classification as a similarity measure in k-nearest neighbor classification [38], which determines the class of a ROI by a majority vote of its k nearest neighbors for a user-specified k.

4.3.2. Results

We performed **leave-one-out** classification experiments [39] on a combined dataset of 21 post-T1 weighted (gadolinium enhanced) ROIs and 75 pre-T2 weighted ROIs extracted from 24 post-T1 and 96 pre-T2 slices of the brain of a single mouse afflicted with a large intracranial neoplasm.

Average kNN accuracies on the combined and individual T1 and T2 datasets for values of k ranging from 1 to 6 are shown in Fig. 6. These experiments are shown for different databases and they clearly and consistently illustrate the out-performance of Joint-ViVo with respect to GLA, InfoLoss and Boosted Weighting for almost all the neighbor size k and the test sets, with only few iterations ($t < 50$ in practice). The out-performance of our visual vocabulary $\mathcal{V}$ comes essentially from the joint learning of the visual vocabulary $\mathcal{V}$ visual words weighting vector $\mathbf{w}$; in almost all cases, 50 iterations was sufficient in order to improve the performance of the $\mathcal{V}$, and few more iterations ($T = 80$) for the other cases. On the one hand, this corroborates the fact that the supervised learned vocabulary and weights provide state of the art performances, and on the other hand, their performances can be consistently improved by jointly learning of visual words and their weights.

5. Conclusion

In this paper, we have explored the use of bag-of-features for tissue classification problems in medical imaging. We commenced by reviewing some of the properties of visual vocabulary and its relationship with the visual words' weighting. This analysis relied on the joint learning of visual vocabulary and the weighting factors, just like the joint learning of features and classifiers in Auto-Context [25, 26]. We proposed a new approach to learn the visual vocabulary and its weights design based on a iterative formed from the generative and discriminative approaches. The main idea is to introduce a discriminative model with a visual words selection criterion to select the most discriminative visual words in each iteration. Two of the most important properties of Joint-ViVo are that the weighting factor is learned based on the vocabulary and that vocabulary is selected based on the learned weighting factors.

We have explored three tissue classification applications of the Joint-ViVo. The first of these is brain tissue types identification in MRI, the second is breast tissue density classification in mammorgram, and the third one is the lung tissue classification in HRCT axial images. In our experiments, we employed three medical image data sets for tissue classification problems and confirmed that the joint learning of visual vocabulary and weighting factors greatly improved the classification performance. One more application is also explored on recognition of natural scene categories. We compared our Joint-ViVo approach with stat-of-the-art visual vocabulary learning and visual words weighting approaches. Our approach greatly outperformed both these approaches when their classification performance was comparable. The novel developed Joint-ViVo algorithm is proven to outperform alternatives in terms of their ability to learn and weight the visual words for bag-of-features method. Moreover, Joint-ViVo algorithm can also be used to bag-of-features based bioinformatics [44, 45, 46, 47, 48, 49], medical imaging [15, 50, 51, 52], biometrics [53, 54, 55, 56, 57, 58] and computer vision [14, 59, 60, 61].

Acknowledgments

The study was supported by grants from Shanghai Key Laboratory of Intelligent Information Processing, China (Grant No. IIPL-2011-003) and Key Laboratory of High Performance Computing and Stochastic Information Processing, Ministry of Education of China (Grant No. HS201107).

References

[1] Y. Yin, M. Adel, S. Bourennane, Retinal vessel segmentation using a probabilistic tracking method, PATTERN RECOGNITION 45 (4) (2012) 1235–1244. doi:{10.1016/j.patcog.2011.09.019}.

[2] Y. Wang, D. Tao, X. Gao, X. Li, B. Wang, Mammographic mass segmentation: Embedding multiple features in vector-valued level set in ambiguous regions, PATTERN RECOGNITION 44 (9, SI) (2011) 1903–1915, 13th International Conference on Computer Analysis of Images and Patterns, Munster, GERMANY, SEP 02-04, 2009. doi:{10.1016/j.patcog.2010.08.002}.

[3] B. Zhang, X. Wu, J. You, Q. Li, F. Karray, Detection of microaneurysms using multi-scale correlation coefficients, PATTERN RECOGNITION 43 (6) (2010) 2237–2248. doi:{10.1016/j.patcog.2009.12.017}.

[4] D. W. Shattuck, S. R. S, K. A. Schaper, D. A. Rottenberg, R. M. Leahy, Magnetic resonance image tissue classification using a partial volume model, NeuroImage 13 (2001) 856–876.

[5] N. Karssemeijer, Automated classification of mammographic parenchymal pattern, Physics in Medicine and Biology 28 (1998) 365C378.

[6] Chuan Zhou, Heang-Ping Chan, N. Petrick, M. Helvie, M. Goodsitt, B. Sahiner, L. Hadjiiski, Computerized image analysis: Estimation of breast density on mammograms, Medical Physics 28 (6) (2001) 1056–69.

[7] A. Bosch, X. Munoz, A. Oliver, J. Marti, Modeling and classifying breast tissue density in mammograms, Vol. 2, New York, NY, United states, 2006, pp. 1552 – 1558.

[8] T. J. Franks, J. R. Galvin, A. A. Frazier, The use and impact of hrct in diffuse lung disease, Current Diagnostic Pathology 10 (2004) 279–290.

[9] G. Sergiacomi, C. Cicci, L. Boi, L. Velari, S. Crusco, A. Orlacchio, G. Simonetti, The use and impact of hrct in diffuse lung disease, Current Diagnostic Pathology 10 (2004) 279–290.

[10] N. Kato, M. Fukui, T. Isozaki, Bag-of-features approach for improvement of lung tissue classification in diffuse lung disease, in: Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 7260, 2009, p. 72600C (10 pp.).

[11] C.-L. Tien, Y.-R. Lyu, S.-S. Jyu, Surface flatness of optical thin films evaluated by gray level co-occurrence matrix and entropy, Applied Surface Science 254 (15) (2008) 4762 – 4767.

[12] H. Jegou, M. Douze, C. Schmid, Improving bag-of-features for large scale image search, International Journal of Computer Vision 87 (3) (2010) 316 – 336.

[13] J. Wang, Y. Li, Y. Zhang, C. Wang, H. Xie, G. Chen, X. Gao, Bag-of-Features Based Medical Image Retrieval via Multiple Assignment and Visual Words Weighting, IEEE Transactions on Medical Imaging 30 (11) (2011) 1996 – 2011. doi:{10.1109/TMI.2011.2161673}.

[14] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, N. Tang, Learning context-sensitive similarity by shortest path propagation, PATTERN RECOGNITION 44 (10-11, SI) (2011) 2367–2374. doi:{10.1016/j.patcog.2011.02.007}.

[15] Jingyan Wang, Yongping Li, Ying Zhang, Honglan Xie, Chao Wang, Boosted learning of visual word weighting factors for bag-of-features based

medical image retrieval, in: Nenghai Yu, Yujin Zhang (Eds.), Proceedings of the Sixth International Conference on Image and Graphics (ICIG 2011), Nat. Natural Sci. Found. China, 2011, pp. 1035–40, 2011 Sixth International Conference on Image and Graphics (ICIG 2011), 12-15 Aug. 2011, Hefei, Anhui, China.

[16] M. Barnathan, J. Zhang, E. Miranda, V. Megalooikonomou, S. Faro, H. Hensley, L. Del Valle, K. Khalili, J. Gordon, F. B. Mohamed, A texture-based methodology for identifying tissue type in magnetic resonance images, Paris, France, 2008, pp. 464 – 467.

[17] N. Bassiou, C. Kotropoulos, Rplsa: A novel updating scheme for probabilistic latent semantic analysis, Computer Speech and Language 25 (4) (2011) 741 – 760.

[18] H. Cai, F. Yan, K. Mikolajczyk, Learning weights for codebook in image classification and retrieval, San Francisco, CA, United states, 2010, pp. 2320 – 2327.

[19] A. M. Bagirov, J. Ugon, D. Webb, Fast modified global k-means algorithm for incremental cluster construction, PATTERN RECOGNITION 44 (4) (2011) 866–876. doi:{10.1016/j.patcog.2010.10.018}.

[20] R. C. de Amorim, B. Mirkin, Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering, PATTERN RECOGNITION 45 (3) (2012) 1061–1075. doi:{10.1016/j.patcog.2011.08.012}.

[21] F. Jurie, B. Triggs, Creating efficient codebooks for visual recognition, Vol. I, Beijing, China, 2005, pp. 604 – 610.

[22] J. C. Van Gemert, J.-M. Geusebroek, C. J. Veenman, A. W. M. Smeulders, Kernel codebooks for scene categorization, Vol. 5304 LNCS, Marseille, France, 2008, pp. 696 – 709.

[23] S. Lazebnik, M. Raginsky, Supervised learning of quantizer codebooks by information loss minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (7) (2009) 1294 – 1309.

[24] M. Lan, C. L. Tan, J. Su, Y. Lu, Supervised and traditional term weighting methods for automatic text categorization, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (4) (2009) 721 – 735.

[25] Z. Tu, Auto-context and its application to high-level vision tasks, Anchorage, AK, United states, 2008.

[26] Z. Tu, X. Bai, Auto-context and its application to high-level vision tasks and 3d brain image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1744 – 1757.

[27] Congyan Lang, Bin Cheng, Songhe Feng, Xiaotong Yuan, Supervised Sparse Patch Coding Towards Misalignment-Robust Face Recognition, in: Nenghai Yu, Yujin Zhang (Eds.), Proceedings of the Sixth International Conference on Image and Graphics (ICIG 2011), Nat. Natural Sci. Found. China, pp. 599–604, 2011 Sixth International Conference on Image and Graphics (ICIG 2011), 12-15 Aug. 2011, Hefei, Anhui, China.

[28] Y. Yang, W. Liu, L. Zhang, Study on improved scale invariant feature transform matching algorithm, Vol. 1, Beijing, China, 2010, pp. 398 – 401.

[29] J. C. Van Gemert, C. J. Veenman, A. W. Smeulders, J.-M. Geusebroek, Visual word ambiguity, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (7) (2010) 1271 – 1283.

[30] J. C. Caicedo, A. Cruz, F. A. Gonzalez, Histopathology image classification using bag of features and kernel functions, Vol. 5651 LNAI, Verona, Italy, 2009, pp. 126 – 135.

[31] M. Varma, A. Zisserman, A statistical approach to material classification using image patch exemplars, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (11) (2009) 2032 – 2047.

[32] V. Athitsos, J. Alon, S. Sclaroff, G. Kollios, Boostmap: An embedding method for efficient nearest neighbor retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (1) (2008) 89 – 104.

[33] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. C. H. Hoi, M. Satyanarayanan, A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 32 (1) (2010) 30–44.

[34] Y. Liu, X.-L. Wang, H.-Y. Wang, H. Zha, H. Qin, Learning robust similarity measures for 3d partial shape retrieval, International Journal of Computer Vision 89 (2-3) (2010) 408 – 431.

[35] X. Lin, T. Qiu, F. Morain-Nicolier, S. Ruan, A topology preserving non-rigid registration algorithm with integration shape knowledge to segment brain subcortical structures from MRI images, PATTERN RECOGNITION 43 (7) (2010) 2418–2427. doi:{10.1016/j.patcog.2010.01.012}.

[36] J. SUCKLING, J. PARKER, D. DANCE, S. ASTLEY, I. HUTT, C. BOGGIS, I. RICKETTS, E. STAMATAKIS, N. CERNEAZ, S. KOK, P. TAYLOR, D. BETAL, J. SAVAGE, THE MAMMOGRAPHIC IMAGE-ANALYSIS SOCIETY DIGITAL MAMMOGRAM DATABASE, in: Gale, AG and Astley, SM and Dance, DR and Cairns, AY (Ed.), DIGITAL MAMMOGRAPHY, Vol. 1069 of INTERNATIONAL CONGRESS SERIES, 1994, pp. 375–378.

[37] K. De Brabanter, P. Karsmakers, J. De Brabanter, J. Suykens, B. De Moor, Confidence bands for least squares support vector machine classifiers: A regression approach, Pattern Recognition (2011) 2280–7.

[38] C.-T. Chang, J. Z. C. Lai, M. D. Jeng, Fast agglomerative clustering using information of k-nearest neighbors, PATTERN RECOGNITION 43 (12) (2010) 3958–3968. doi:{10.1016/j.patcog.2010.06.021}.

[39] J. Wu, J. Mei, S. Wen, S. Liao, J. Chen, Y. Shen, A self-adaptive genetic algorithm-artificial neural network algorithm with leave-one-out cross validation for descriptor selection in qsar study, Journal of Computational Chemistry 31 (10) (2010) 1956 – 1968.

[40] J.-H. Hong, S.-B. Cho, A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification, Vol. 71, P.O. Box 211, Amsterdam, 1000 AE, Netherlands, 2008, pp. 3275 – 3281.

[41] L. Blot, R. Z. Y, Background texture extraction for the classification of mammographic parenchymal patterns, in: In MIUA, 2001, pp. 145–148.

[42] A. Oliver, J. Freixenet, A. Bosch, D. Raba, R. Zwiggelaar, Automatic classification of breast tissue (2005).

[43] C.-Q. Chen, An enhanced generalized lloyd algorithm, IEEE Signal Processing Letters 11 (2 PART II) (2004) 167 – 170.

[44] J. Wang, Y. Li, SEQUENTIAL LINEAR NEIGHBORHOOD PROPAGATION FOR SEMI-SUPERVISED PROTEIN FUNCTION PREDICTION, JOURNAL OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY 9 (6) (2011) 663–679. doi:{10.1142/S0219720011005550}.

[45] Jingyan Wang, Yongping Li, Ying Zhang, Ning Tang, Chao Wang, Class Conditional Distance Metric for 3D Protein Structure Classification, in: 2011 5th International Conference on Bioinformatics and Biomedical Engineering, Eng. Medicine Biol. Soc., 2011, p. 4 pp., 2011 5th International Conference on Bioinformatics and Biomedical Engineering, 10-12 May 2011, Wuhan, China.

[46] Jingyan Wang, Yongping Li, Ying Zhang, Jianhua He, Semi-supervised Protein Function Prediction via Sequential Linear Neighborhood Propagation, in: De-Shuang Huang, Yong Gan, P. Premaratne, Kyungsook Han (Eds.), Bio-Inspired Computing and Applications. 7th International Conference on Intelligent Computing, ICIC 2011.Revised Selected Papers,

2011, pp. 435–41, Bio-Inspired Computing and Applications. 7th International Conference on Intelligent Computing, ICIC 2011, 11-14 Aug. 2011, Zhengzhou, China.

[47] J. Wang, X. Gao, Q. Wang, Y. Li, ProDis-ContSHC: learning protein dissimilarity measures and hierarchical context coherently for protein-protein comparison in protein database retrieval, BMC BIOINFORMATICS 13 (7), International Conference on Intelligent Computing (ICIC), Zhengzhou, PEOPLES R CHINA, AUG 11-14, 2011. doi:{10.1186/1471-2105-13-S7-S2}.

[48] J. Wang, Y. Li, Q. Wang, X. You, J. Man, C. Wang, X. Gao, ProClusEnsem: Predicting membrane protein types by fusing different modes of pseudo amino acid composition, COMPUTERS IN BIOLOGY AND MEDICINE 42 (5) (2012) 564–574. doi:{10.1016/j.compbiomed.2012.01.012}.

[49] H. Pei, J. Li, M. Lv, J. Wang, J. Gao, J. Lu, Y. Li, Q. Huang, J. Hu, C. Fan, A graphene-based sensor array for high-precision and adaptive target identification with ensemble aptamers, Journal of the American Chemical Society.

[50] Jingyan Wang, Yongping Li, Ying Zhang, Honglan Xie, Chao Wang, Bag-of-Features Based Classification of Breast Parenchymal Tissue in the Mammogram via Jointly Selecting and Weighting Visual Words, in: Nenghai Yu, Yujin Zhang (Eds.), Proceedings of the Sixth International Conference on Image and Graphics (ICIG 2011), Nat. Natural Sci. Found. China, 2011, pp. 622–7, 2011 Sixth International Conference on Image and Graphics (ICIG 2011), 12-15 Aug. 2011, Hefei, Anhui, China.

[51] J. Wang, Y. Li, E. Marchiori, C. Wang, Iterated large-margin discriminant analysis for feature dimensionality reduction in medical image retrieva International Journal of Biomedical Engineering and Technology 7 (2) (2011) 116 – 134, class conditional neighbouring graph;Dimensionality

reduction;DR;Expectation-maximisation;Iterated large-margin discriminat analysis;Medical image retrieval;.
URL http://dx.doi.org/10.1504/IJBET.2011.043174

[52] J. Wang, I. Almasri, Modeling multiple visual words assignment for bag-of-features based medical image r in: Proceedings of the IASTED International Conference on Computer Graphics and Imaging, CGIM 2012, Crete, Greece, 2012, pp. 217 – 224.
URL http://dx.doi.org/10.2316/P.2012.779-015

[53] J. Wang, Y. Li, P. Liang, G. Zhang, X. Ao, An Effective Multi-Biometrics Solution for Embedded Device, in: 2009 IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS (SMC 2009), VOLS 1-9, IEEE International Conference on Systems Man and Cybernetics Conference Proceedings, IEEE, 2009, pp. 917–922, IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, OCT 11-14, 2009. doi:{10.1109/ICSMC.2009.5346745}.

[54] J. Wang, Y. Li, X. Ao, C. Wang, J. Zhou, MULTI-MODAL BIOMETRIC AUTHENTICATION FUSING IRIS AND PALMPRINT BASED ON GMM, in: 2009 IEEE/SP 15TH WORKSHOP ON STATISTICAL SIGNAL PROCESSING, VOLS 1 AND 2, IEEE; SP, 2009, pp. 349–352, 15th IEEE/SP Workshop on Statistical Signal Processing, Cardiff, WALES, AUG 31-SEP 03, 2009. doi:{10.1109/ICICTA.2009.551}.

[55] Jingyan Wang, Yongping Li, Chao Wang, How to handle missing data in robust multi-biometrics verification, International Journal of Biometrics 3 (2011) 265–83.

[56] X. Zhai, Y. Zhao, J. Wang, Y. Li, Q. Wang, H. Xie, Adaptive SVM Fusion for Robust Multi-Biometrics Verification with Missing Data, in: Zhou, X (Ed.), 2011 INTERNATIONAL CONFERENCE ON ENERGY AND ENVIRONMENTAL SCIENCE-ICEES 2011, Vol. 11 of Energy Procedia, Singapore Inst Electron, 2011, pp. 1006–1012, International Conference

on Energy and Environmental Science (ICEES), Singapore, SINGAPORE, OCT, 2011. doi:{10.1016/j.egypro.2011.10.349}.

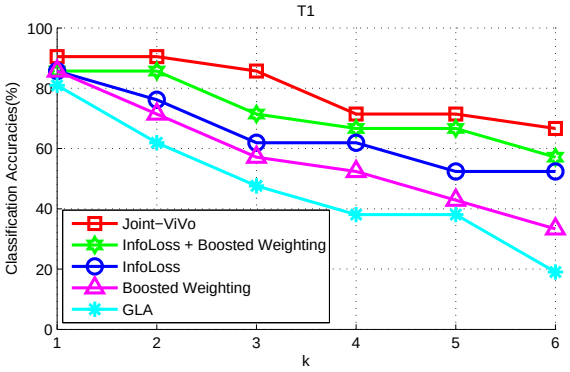
[57] J. Liu, Y. Hou, J. Wang, Y. Li, Q. Wang, J. Man, H. Xie, J. He, Fusing Iris and Palmprint at Image Level for Multi-Biometrics Verification, in: Zeng, Z and Li, Y (Ed.), FOURTH INTERNATIONAL CONFERENCE ON MACHINE VISION (ICMV 2011): COMPUTER VISION AND IMAGE ANALYSIS: PATTERN RECOGNITION AND BASIC TECHNOLOGIES, Vol. 8350 of Proceedings of SPIE, Singapore Inst Elect; Int Assoc Comp Sci & Informat Technol (IACSIT), 2012, 4th International Conference on Machine Vision (ICMV) - Computer Vision and Image Analysis - Pattern Recognition and Basic Technologies, Singapore, SINGAPORE, DEC 09-10, 2011. doi:{10.1117/12.920534}.

[58] J. Wang, Y. Li, Y. Zhang, Y. Huang, Implementing multimodal biometric solutions in embedded systems, Biometrics - Unique and Diverse Applications in Nature.

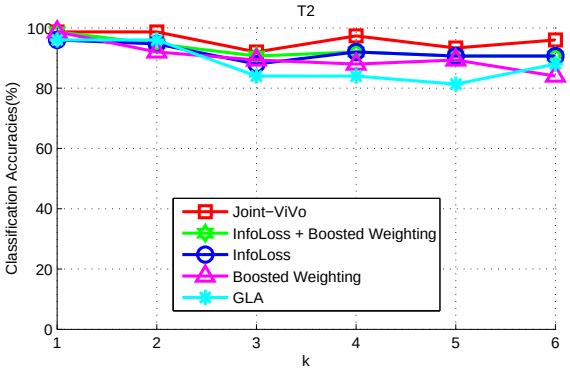
[59] Zhonghua Liu, Jingyan Wang, Yongping Li, Ying Zhang, Chao Wang, Quantized image patches co-occurrence matrix: a new statistical approach for texture classification using image patch exemplars, in: Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 8009, 2011, p. 80092P (5 pp.), Third International Conference on Digital Image Processing (ICDIP 2011), 15-17 April 2011, Chengdu, China.

[60] J. Wang, M. A. Jabbar, Multiple kernel learning for adaptive graph regularized nonnegative matrix factor in: Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications, SPPRA 2012, Crete, Greece, 2012, pp. 115 – 122.
URL http://dx.doi.org/10.2316/P.2012.778-049

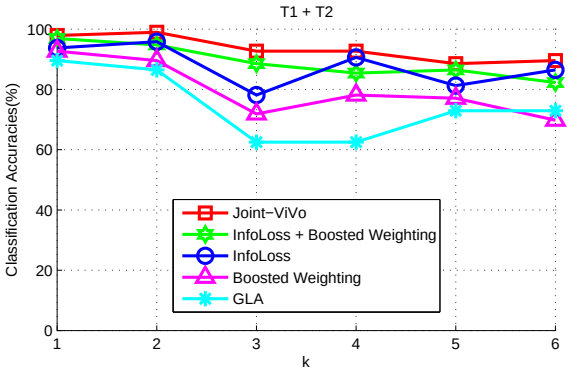
[61] Z. Liu, J. Wang, J. Man, Y. Li, X. You, C. Wang, Self-adaptive Local Fisher Discriminant Analysis for semi-supervised image recognition, International Journal of Biometrics.



(a) T1



(b) T2



(c) T1 + T2

Figure 6: Average kNN classification accuracies for the T1 (a), T2 (b) and combined datasets (c).